



**(α) ΔΙΠΜΣ ΣΤΙΣ ΕΦΑΡΜΟΣΜΕΝΕΣ ΜΑΘΗΜΑΤΙΚΕΣ ΕΠΙΣΤΗΜΕΣ**  
**(β) ΔΙΠΜΣ ΣΤΗΝ ΜΑΘΗΜΑΤΙΚΗ ΠΡΟΤΥΠΟΠΟΙΗΣΗ ΣΕ ΣΥΓΧΡΟΝΕΣ**  
**ΤΕΧΝΟΛΟΓΙΕΣ ΚΑΙ ΣΤΗΝ ΟΙΚΟΝΟΜΙΑ**

**ΜΑΘΗΜΑ:** *Μπεϋζιανή Στατιστική και MCMC*

**ΔΙΔΑΣΚΩΝ:** *ΔΗΜΗΤΡΗΣ ΦΟΥΣΚΑΚΗΣ (τηλ: 210 7721702 – email:*

*fouskakis@math.ntua.gr)*

## **ΕΡΓΑΣΙΑ**

### **Άσκηση 1**

Τα παρακάτω δεδομένα (*data*)  $y$  μας δίνουν τη μέση μηνιαία βροχόπτωση, όπως καταγράφηκε από το Σύστημα Πληροφοριών για τη Διαχείριση Υδάτινων Πόρων της Πολιτείας του Σάο Πάολο για 56 έτη (1947-2003) τον μήνα Νοέμβριο

0.2 3.5 2.8 3.7 8.7 6.9 7.4 0.8 4.8 2.5 2.9 3.1 4.0 5.0 3.8 3.5 5.4 3.3 2.9 1.7 7.3 2.9 4.6  
1.1 1.4 3.9 6.2 4.1 10.8 3.8 7.3 1.8 6.7 3.5 3.2 5.2 2.8 5.2 5.4 2.2 9.9 2.1 4.7 5.5 2.6  
4.1 5.4 5.5 2.1 1.9 8.8 1.3 24.1 5.4 6.2 2.9

Θεωρήστε ότι τα παραπάνω δεδομένα προέρχονται από τη Γάμμα κατανομή με σ.π.π.

$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$ ,  $x > 0$  και παραμέτρους  $\alpha > 0$  και  $\beta > 0$ . Θεωρήστε την παράμετρο

$\alpha$  γνωστή και ίση με 2.4 και την παράμετρο  $\beta$  άγνωστη.

- Υπολογίστε αναλυτικά την ύστερη κατανομή της παραμέτρου  $\beta$ , χρησιμοποιώντας μια μη πληροφοριακή συζυγή πρότερη. Εξηγήστε τον τρόπο με τον οποίο θα βρείτε τη συζυγή πρότερη και σχολιάστε πως θα γίνει μη πληροφοριακή. Περιγράψτε κατάλληλα, υπολογίζοντας μέτρα θέσης και μεταβλητότητας, την ύστερη κατανομή της παραμέτρου  $\beta$ . Σε ένα διάγραμμα αναπαραστήστε την πιθανοφάνεια, την πρότερη και την ύστερη κατανομή της παραμέτρου  $\beta$  και σχολιάστε.

- ii. Με χρήση της ύστερης κατανομής του ερωτήματος i., υπολογίστε αναλυτικά την ύστερη προβλεπτική κατανομή της 57<sup>ης</sup> παρατήρησης και εκτιμήστε την 57<sup>η</sup> παρατήρηση από τον μέσο της<sup>1</sup>. Συγκρίνετε την τιμή αυτή με την τιμή που θα είχατε αν είχατε κάνει την πρόβλεψη με χρήση της κλασικής στατιστικής.
- iii. Θεωρώντας εκ των προτέρων τις υποθέσεις  $H_0: \beta = 1$  έναντι της  $H_1: \beta \neq 1$  ισοπίθανες, υπολογίστε αναλυτικά τον παράγοντα Μπέυζ<sup>2</sup> της μηδενικής υπόθεσης έναντι της εναλλακτικής και συνδέστε τον με τον ύστερο λόγο των συμπληρωματικών πιθανοτήτων (*Posterior Odds* –  $PO_{01}$ ). Κάτω από την  $H_1$  χρησιμοποιείτε την πρότερη κατανομή του ερωτήματος i. για την παράμετρο  $\beta$ . Ποιο είναι το τελικό σας συμπέρασμα για τον έλεγχο υπόθεσης; (Υπόδειξη: Υλοποιήστε τους υπολογισμούς σας σε λογαριθμική κλίμακα).

## Άσκηση 2

Θεωρήστε τα δεδομένα της Άσκησης 1 και υποθέστε τώρα πως οι παράμετροι της Γάμμα κατανομής είναι και οι δύο άγνωστες. Χρησιμοποιώντας την R δημιουργήστε τον δικό σας αλγόριθμο *Gibbs* για να προσομοιώσετε τιμές από την ύστερη κατανομή των  $\alpha$  και  $\beta$ <sup>3</sup>. Σε περίπτωση που οι πλήρους δέσμευσης ύστερες κατανομές  $\alpha | (\beta, y)$  και  $\beta | (\alpha, y)$  δεν υπολογίζονται σε κλειστή μορφή χρησιμοποιείτε μέσα στον αλγόριθμο *Gibbs* τον αλγόριθμο *Metropolis-Hastings* με Κανονική κατανομή εισήγησης για να προσομοιώσετε τιμές από αυτές. Δώστε ιστογράμματα και περιγραφικούς πίνακες των προσομοιωμένων τιμών. Δώστε επίσης *trace plots*, *ergodic mean plots* και διαγράμματα αυτοσυσχετίσεων και συζητήστε για την σύγκλιση του αλγορίθμου σας καθώς και για την επιλογή του *burn-in* και του συνολικού αριθμού επαναλήψεων.

## Άσκηση 3

Από όποια πηγή επιθυμείτε βρείτε κάποια δεδομένα αντίστοιχα με αυτά του Παραδείγματος 3 με τα ποντίκια που κάναμε στην τάξη στο τελευταίο σετ

<sup>1</sup> **Υπόδειξη:** Θα σας βοηθήσει να κοιτάξετε την *Generalized Beta Prime* κατανομή.

<sup>2</sup> **Υπόδειξη:** Για τον υπολογισμό του παράγοντα Μπέυζ θα πρέπει να υπολογίσετε τις περιθώριες πιθανοφάνειες κάτω από τις δύο υποθέσεις. Κάτω από την  $H_1$  δουλέψετε με τη δειγματοληπτική κατανομή του επαρκούς στατιστικού για την παράμετρο  $\beta$ .

<sup>3</sup> **Υπόδειξη:** Χρησιμοποιείτε μη πληροφοριακές Γάμμα κατανομές ως πρότερες για τις παραμέτρους  $\alpha$  και  $\beta$  (δηλαδή Γάμμα(0.001, 0.001)). Ξεκινήστε με “ρεαλιστικές” αρχικές τιμές χρησιμοποιώντας π.χ. την εντολή `fitdistr(data, "gamma")` της R, αφού πρώτα εξηγήσετε τι κάνει.

διαφανειών. Θεωρήστε δηλαδή ότι η μεταβλητή απόκρισης ακολουθεί τη Διωνυμική κατανομή και η επεξηγηματική σας μεταβλητή είναι κατηγορική με 3 ή περισσότερες κατηγορίες. Για μία μόνο από τις κατηγορίες της επεξηγηματικής σας μεταβλητής υποθέστε ότι το ίδιο πείραμα είχε επαναληφθεί στο παρελθόν, υποθέστε κάποια δεδομένα για αυτό και εισάγετε κατάλληλα την πληροφορία αυτή στην πρότερη κατανομή της τωρινής σας ανάλυσης. Με την βοήθεια του *WinBugs*, βγάλτε κατάλληλα συμπεράσματα και συγκρίνετε τα αποτελέσματα σας με αυτά από την κλασική στατιστική.

#### Άσκηση 4

Τα δεδομένα στον παρακάτω σύνδεσμο

<http://www.math.ntua.gr/~fouskakis//Bayesian/Assign/data.csv>

μας δίνουν την εβδομαδιαία δαπάνη για τρόφιμα σε δολάρια (*food*), το εβδομαδιαίο εισόδημα σε δολάρια (*income*) και τον αριθμό ατόμων (*persons*) 38 νοικοκυριών στις Η.Π.Α. Αρχικά δημιουργήστε μια νέα μεταβλητή (*perc\_food*) η οποία θα καταγράφει το ποσοστό του εισοδήματος των νοικοκυριών που δαπανάται για τρόφιμα την εβδομάδα, με τιμές στο  $(0, 1)$ . Θεωρήστε ότι αυτή η νέα μεταβλητή ακολουθεί τη Βήτα κατανομή με άγνωστες παραμέτρους  $\alpha > 0$  και  $\beta > 0$ . Γράψτε αρχικά τις παραμέτρους της Βήτα κατανομής, ως συναρτήσεις της μέσης της τιμής  $\mu$  και της παραμέτρου  $\gamma = \alpha + \beta$  και σχολιάστε πως η παράμετρος  $\gamma$  επηρεάζει την διασπορά της Βήτα κατανομής.

- i. Χρησιμοποιώντας το *WinBugs* προσαρμόστε ένα Μπεϋζιανό μοντέλο Βήτα παλινδρόμησης, με μεταβλητή απόκρισης την *perc\_food*, και επεξηγηματικές μεταβλητές τις *income* και *persons*. Θεωρήστε πως η μέση τιμή μπορεί να γραφτεί, με χρήση της *logit* συνδετικής συνάρτησης, ως γραμμική συνάρτηση των επεξηγηματικών μεταβλητών, ενώ υποθέστε ομοσκεδαστικότητα (κοινή παράμετρο  $\gamma$  για κάθε νοικοκυριό). Χρησιμοποιείστε κατάλληλες μη πληροφοριακές πρότερες κατανομές για τους συντελεστές του μοντέλου, όπως και για την παράμετρο  $\gamma$ , και εκτιμήστε τις μέσες τιμές των ύστερων κατανομών των εν λόγω συντελεστών και ερμηνεύστε τις κατάλληλα με χρήση *odds* και *odds ratios*.
- ii. Χρησιμοποιώντας το *WinBugs* προσαρμόστε ένα Μπεϋζιανό μοντέλο Βήτα παλινδρόμησης, με μεταβλητή απόκρισης την *perc\_food*, και

επεξηγηματικές μεταβλητές τις *income* και *persons*. Θεωρήστε πως η μέση τιμή μπορεί να γραφτεί, με χρήση της *probit* συνδυαστικής συνάρτησης, ως γραμμική συνάρτηση των επεξηγηματικών μεταβλητών, ενώ υποθέστε ομοσκεδαστικότητα (κοινή παράμετρο  $\gamma$  για κάθε νοικοκυριό). Χρησιμοποιείστε κατάλληλες μη πληροφοριακές πρότερες κατανομές για τους συντελεστές του μοντέλου, όπως και για την παράμετρο  $\gamma$ , και εκτιμήστε τις μέσες τιμές των ύστερων κατανομών των εν λόγω συντελεστών και ερμηνεύστε τις κατάλληλα. Για την ερμηνεία των συντελεστών του εν λόγω μοντέλου, πολλαπλασιάστε τους ύστερους μέσους των συντελεστών με την τιμή 1.6 και εν συνεχεία, προσεγγιστικά, η ερμηνεία και πάλι μπορεί να γίνει όπως και στο *logit* μοντέλο, ως προς *odds* και *odds ratios*.

- iii. Χρησιμοποιώντας το εργαλείο *DIC* στο *WinBUGS* αποφανθείτε ποιο από τα δύο μοντέλα (ερωτήματα i. και ii.) τελικά θα επιλέγατε<sup>4</sup>.

## Οδηγίες

- Η εργασία θα πρέπει να παραδοθεί ηλεκτρονικά στο email μου, [fouskakis@math.ntua.gr](mailto:fouskakis@math.ntua.gr), μέχρι την Πέμπτη 27 Ιουνίου 2024 στις 13:00μμ. Καμιά εργασία δεν θα γίνει δεκτή μετά την ώρα αυτή.
- Η εργασία που θα παραδώσετε πρέπει να είναι σε *pdf* μορφή αφού πρώτα τη γράψετε υποχρεωτικά σε *Latex*. Παρακαλώ χρησιμοποιήστε τον ακόλουθο τίτλο στο *pdf* αρχείο σας: Surname-Name.pdf, όπου Surname είναι το επώνυμό σας (με λατινικούς χαρακτήρες) και Name το όνομα σας (με λατινικούς χαρακτήρες). Π.χ. αν παρέδιδα εγώ την εργασία θα την ονόμαζα ως εξής: Fouskakis-Dimitris.pdf.
- Παρακαλώ χρησιμοποιήστε ένα εξώφυλλο στο *pdf* αρχείο σας, στο οποίο να υπάρχει κατάλληλος τίτλος και να αναγράφεται υποχρεωτικά το ονοματεπώνυμο σας, το μεταπτυχιακό (ή διδακτορικό) πρόγραμμα που παρακολουθείτε καθώς και το email σας και ο αριθμός μητρώου σας.
- Θα πρέπει να αποστείλετε ένα μόνο αρχείο. Η εργασία θα πρέπει να περιλαμβάνει τους κώδικες της R και του *WinBUGS*, όχι σε παράρτημα

---

<sup>4</sup>Υπόδειξη: Αν σε κάποιο μοντέλο δεν εμφανίζεται η ένδειξη *DIC*, χρησιμοποιήστε και άλλες επαναλήψεις για *burn-in*.

αλλά στην απάντηση του κάθε ερωτήματος, με πλήρη επεξήγηση, γραφήματα και πλήρη περιγραφή των αποτελεσμάτων.

- Θα δοθεί ιδιαίτερη σημασία στην παρουσίαση της εργασίας. Η εργασία πρέπει να είναι κατανοητή και να περιγράφει ο,τιδήποτε χρησιμοποιήσατε πειστικά για κάποιον που δεν γνωρίζει πάρα πολλά για το αντικείμενο.
- Την **Τετάρτη 17 Ιουλίου 2024** θα υπάρξει παρουσίαση της εργασίας στο *PC-LAB* του Τομέα Μαθηματικών. **Οι παρουσιάσεις θα ξεκινήσουν στις 10.00πμ.** Όλοι οι φοιτητές, που παρέδωσαν εργασία, εκείνη την μέρα θα πρέπει να παρευρίσκονται σε όλες τις παρουσιάσεις και ο καθένας μεμονωμένα θα παρουσιάσει μέρος της εργασίας του. **Η παρουσία όλων εκείνη την μέρα είναι υποχρεωτική. Σε περίπτωση απουσίας η εργασία δεν θα βαθμολογηθεί.** Για την παρουσίαση αρκεί να έχετε μαζί σας σε ένα *usb-stick* την εργασία σας σε *pdf* μορφή. Μετά το πέρας των παρουσιάσεων θα καλείται στο γραφείο μου ο κάθε φοιτητής μεμονωμένα, για τα τελικά σχόλια και την ανακοίνωση του βαθμού του.

**Εύχομαι Επιτυχία**